

WHAT IS CLAIMED IS:

Sub
A4

1. A document information search apparatus for searching document information on the basis of a search request transmitted through a network and responding, wherein:

a search condition designating unit which designates a file as a search condition and transmits contents of said designated file via the network is provided for a search requesting source; and

- 10 a document search unit which forms a keyword from the file contents transmitted from said search condition designating unit and searches similar documents from a database is provided on a search side.

- 15 2. An apparatus according to claim 1, wherein said search condition designating unit transmits a head file portion of the designated file contents.

- 20 3. An apparatus according to claim 1, wherein said search condition designating unit allows an HTML file and an Excel file to be included in the file which is designated as said search condition.

- 25 4. An apparatus according to claim 1, wherein index information describing a list of important words extracted from search target documents is stored every document in said database,

10951223-041604

and said document search unit on the search side comprises:

5 a text extraction processing unit which extracts a text document from the file contents received in response to the search request;

a morpheme analyzing unit which extracts nouns by a morpheme analysis of said text document;

10 a keyword forming unit which extracts important words from said nouns and forms a keyword in which said important words are coupled by OR; and

15 a search executing unit which searches similar documents by searching the search database by said keyword and notifies the search requesting source of a search result.

20 5. An apparatus according to claim 4, wherein said keyword forming unit counts the number of times of appearance showing in which documents in the index of each of the search documents stored in said document database each of said nouns appears, selects a predetermined number of upper words each having the number of times of appearance in a predetermined range, and forms the keyword.

25 6. An apparatus according to claim 5, wherein in the case where the number of documents in the index is assumed to be (N), said keyword forming unit selects

upper ten words each having the number (H) of times of appearance in a range where $2N/3 \geq H \geq 1$ and forms the keyword.

- 5 7. An apparatus according to claim 5, wherein said keyword forming unit allows property information extracted from the file received in response to the search request to be included in said keyword, thereby allowing the similar documents to be searched.

10

8. An apparatus according to claim 7, wherein said property information includes a writer of the file received in response to the search request, a document title, and the like.

15

9. An apparatus according to claim 1, wherein said search condition designating unit of said search requesting source is provided by a WWW browser of a client, transmits the contents of the file designated by a search request picture plane of said WWW browser to a search machine of a WWW server through the network, and sends said file contents to said document search unit.

20

- 25 10. A document information search apparatus comprising:

a database in which index information describing a

list of important words extracted from search target documents has been stored every document;

5 a text extraction processing unit which extracts a text document from contents of a file received in response to a search request from a network in which a document file has been designated as a search condition;

a morpheme analyzing unit which extracts nouns by a morpheme analysis of said text document;

10 a keyword forming unit which extracts important words from said nouns and forms a keyword in which said important words are coupled by OR; and

15 a search executing unit which searches similar documents by searching said database by said keyword and notifies a requesting source of a search result.

20 11. An apparatus according to claim 10, wherein said keyword forming unit counts the number of times of appearance showing in which documents in the index of each of the search documents stored in said document database each of said nouns appears, selects a predetermined number of upper words each having the number of times of appearance in a predetermined range, and forms the keyword.

25 12. An apparatus according to claim 10, wherein property information extracted from the search target

documents is stored in said database together with the index information, said keyword forming unit allows said property information extracted from the file received in response to the search request to be
5 included in said keyword, thereby allowing the similar documents to be searched.

13. A document information search method of searching document information on the basis of a search request
10 transmitted via a network and responding, comprising the steps of:

storing index information describing a list of important words extracted from search target documents every document into a database;

15 in the case where a file is designated as a search condition on a search requesting source, transmitting contents of the designated file to a server together with the search request through the network; and

on a search side, extracting a text document from
20 the file contents received in response to the search request, extracting nouns by a morpheme analysis of said text document, extracting important words from said nouns, forming a keyword in which said important words are coupled by OR, searching similar documents by
25 searching said database by said keyword, and responding a search result.

14. A method according to claim 13, wherein as a
formation of said keyword, the number of times of
appearance showing in which documents in the index of
each of the documents stored in said database each of
5 said nouns appears is counted, a predetermined number
of upper words each having the number of times of
appearance in a predetermined range are selected, and
the keyword is formed.

10 15. A method according to claim 14, wherein property
information extracted from the file received in
response to the search request is included in said
keyword and the similar documents are searched.

15 16. A computer-readable recording medium in which a
search program has been stored, wherein said program
comprises the steps of:

receiving a search request in which a document
file has been designated as a search condition;

20 extracting a text document from contents of the
file received in response to the search request;

extracting nouns by a morpheme analysis of said
text document;

25 extracting important words from said nouns and
forming a keyword in which said important words are
coupled by OR; and

searching similar documents by searching a

database by said keyword and notifying a requesting source of a search result.

17. A medium according to claim 16, wherein in the
5 step which forms the keyword of said search program,
the number of times of appearance showing in which
documents in the index of each of the documents stored
in said database each of said nouns appears is counted,
a predetermined number of upper words each having the
10 number of times of appearance in a predetermined range
are selected, and the keyword is formed.

18. A medium according to claim 16, wherein said
search program further comprises a step which allows
15 property information extracted from the file received
in response to the search request to be included in
said keyword and searches the similar documents.